

Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Tomáš Krajňák

Rozdělení výši škod z operačního rizika

Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Lucie Mazurová, Ph.D.

Studijní program: Matematika

Studijní obor: Finanční matematika

Praha 2012

Na tomto mieste by som sa chcel poďakovať vedúcej mojej bakalárskej práce RNDr. Lucii Mazurovej, Ph.D., za jej odbornú pomoc, cenné rady, pripomienky, poskytnutie literatúry a predovšetkým za jej čas strávený pri konzultáciách. Ďalej by som sa chcel poďakovať mojej rodine, ktorá ma podporuje v štúdiu.

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona.

V Praze dne 2. 8. 2012

Tomáš Krajňák

Název práce: Rozdělení výší škod z operačního rizika

Autor: Tomáš Krajňák

Katedra / Ústav: Katedra pravděpodobnosti a matematické statistiky

Vedoucí bakalářské práce: RNDr. Lucie Mazurová, Ph.D.

Abstrakt: Operačné riziko sa v posledných rokoch stalo dôležitou súčasťou bánk, poisťovní a finančných inštitúcií. Predložená práca sa zaoberá rozdeleniami, ktoré najlepšie popisujú výšku škôd z operačného rizika a popisuje ich základné vlastnosti. Presnejšie sa zaoberá $g - h$ rozdelením, jeho vlastnosťami, momentmi, odhadmi parametrov a chovaním chvostov. V práci je taktiež popísaná metóda na odhadovanie pravdepodobností prekročenia vysokých hraníc, metóda POT (Peaks Over Threshold). V závere práce je spracovaný postup na odhadovanie kvantilov $g - h$ rozdelenia pomocou POT metódy a je tam uvedený simulačný príklad, v ktorom sú porovnané hodnoty kvantilov odhadovaných pomocou metódy POT s kvantilmi pre $g - h$ rozdelenie.

Klíčová slova: Operačné riziko, $g - h$ rozdelenie, POT metóda

Title: Operational risk loss distributions

Author: Tomáš Krajňák

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Lucie Mazurová, Ph.D.

Abstract: Operational risk in recent years has become an important part of banks, insurance companies and financial institutions. The proposed work deals with the distributions that best fit the loss severity from the operational risk and also describe their basic properties. Specifically, deals with the $g-h$ distribution, its properties, moments, parameter estimations and tail behavior. There is also another method for high threshold estimation described in this text, the POT (Peaks over threshold). In conclusion, there is the procedure for estimating quantiles of $g-h$ distribution by POT method presented including simulation example in which there are quantile values estimated using the POT method compared to the $g-h$ distribution quantiles.

Keywords: operational risk, $g-h$ distribution, POT method

Obsah

Úvod.....	1
1. Charakteristika operačného rizika.....	2
2. Parametrické rozdelenia	4
2.1 LDA prístup	4
2.2 Parametrické rozdelenia pre veľkosti škôd	4
2.2.1 Exponenciálne rozdelenie	6
2.2.2 Weibullovo rozdelenie	7
2.2.3 Gama rozdelenie.....	8
2.2.4 Logaritnicko – normálne rozdelenie	9
2.2.5 Logaritnicko- logistické rozdelenie	10
2.2.6 Zovšeobecnené Paretovo rozdelenie	11
3. $G-h$ rozdelenie a jeho vlastnosti	13
3.1 Základná definícia.....	13
3.2 Odhad parametrov $g - h$ rozdelenia	15
3.3 Momenty $g - h$ rozdelenia.....	17
3.3.1 Stredná hodnota.....	18
3.3.2 Rozptyl	18
3.3.3 Šikmosť	19
3.3.4 Špicatosť	19
3.4 Chovanie chvostov	20
3.4.1 Subexponenciálne rozdelenia.....	22
4. Teória extrémnych hodnôt	24
4.1 POT metóda	24
4.2 Odhady kvantilov	27

4.2.1 Aplikácia POT metódy.....	28
Záver	32
Literatúra	33

Úvod

V živote každého jednotlivca, spoločnosti čiha každodenne mnoho nástrah – rizík. Možnostiam rizika je obzvlášť vystavená taká práca, pri ktorej sa pracuje s peniazmi. Transakcie s peniazmi sú náplňou rôznych inštitúcií, bánk či poisťovní. Tieto transakcie sú vykonávané ľuďmi alebo pomocou počítačov. Riziko omylov a chýb je pomerne vysoké.

Cieľom bakalárskej práce „Rozdělení výší škod z operačního rizika“ budú rozdelenia pravdepodobnosti navrhované pre modelovanie škôd z operačného rizika a zameriame sa hlavne na $g - h$ rozdelenie. Uvádzajú sa tu niektoré matematické prístupy, ako napríklad LDA (Loss distribution approach) alebo POT (Peaks over threshold) metóda. V závere práce je uvedený simulačný príklad pre odhad vysokých kvantilov $g - h$ rozdelenia, čo vlastne uvádza pravdepodobnosť vzniku zriedkavej, ale extrémnej udalosti.

Prvá kapitola bakalárskej práce je zameraná na charakteristiku operačného rizika, opisuje operačné riziko, jeho definíciu a vznik.

Druhá kapitola sa venuje LDA prístupu a sú tam popísané parametrické rozdelenia, ich charakteristiky, význam, grafy a použitie.

Tretia kapitola sa orientuje na $g - h$ rozdelenie a jeho analýzu. Ukazuje odhady parametrov tohto rozdelenia, chovanie chvostov, definuje subexponencialitu rozdelenia a charakteristiky ako stredná hodnota, rozptyl, šikmosť a špicatosť.

Štvrtá kapitola popisuje teóriu extrémnych hodnôt a s ňou spojenú POT (Peaks over threshold) metódu. Poukazuje na aproximáciu pre excesy GPD (Generalized Pareto distribution) rozdelením a ukazuje možnosti odhadu vysokých kvantilov $g - h$ rozdelenia. Je tu taktiež uvedený simulačný príklad.

V záverečnej časti je stručné zhrnutie bakalárskej práce, zhodnotenie vytýčeného cieľa, dôležitosť $g - h$ rozdelenia pri určovaní škôd plynúcich z operačného rizika ako aj jeho význam.

1. Charakteristika operačného rizika

Operačné riziko sa vyskytuje v bankovníctve a vo finančnom sektore už dlhé roky, ale iba pred pár rokmi sa stalo predmetom regulácie, riadenia a diskusií. V minulosti mnoho bánk vykazovalo vysoké straty, ktoré boli spôsobené práve v dôsledku operačného rizika. Niekedy to malo až katastrofálne následky. Niektoré banky dokonca skrachovali kvôli týmto stratám. V porovnaní s tržným a úverovým rizikom venovali banky tomuto riziku iba minimálnu pozornosť a považovali ho veľmi dlho iba za „ostatné riziko“. To malo za následok, že strát bolo čím ďalej tým viac, príčin týchto strát bolo tiež veľké množstvo a neustále ich pribúdalo. Tieto príčiny sa zvyšovali tak, ako sa zvyšovalo zapájanie sa zamestnancov a rôznych systémov (software) do činnosti banky. Outsourcing, rozšírenie internetu, či internetové bankovníctvo na jednej strane zlepšilo prístup k informáciám, ale na strane druhej zvyšovalo šance k pochybeniam, zlyhaniam a finančným únikom a tým k rastu významu operačného rizika. A to bol dobrý dôvod k tomu, aby sa banky začali venovať tejto problematike a vytvorili adekvátny systém pre rozpoznávanie príčin a ochranu pred nimi.

S pojmom operačné riziko sa stretávame vo veľkej miere až pri jeho zaradení do Basel II. To znamená, že banky začali sledovať, zaznamenávať a riadiť operačné riziko približne okolo roku 2004. Tento pojem je ťažko definovateľný aj napriek tomu, že v súčasnej dobe mu je venovaná dostatočne veľká pozornosť. V najširšom alebo všeobecnom poňatí je možné ho definovať ako akékoľvek riziko, ktoré nie je kreditné alebo tržné. Dá sa však chápať ako riziko straty v dôsledku prevádzkových nedostatkov a chýb, respektíve ako riziko, ktoré plyní z operačných činností firmy. Definícia tohto rizika však nie je úplne jednotná na rozdiel od ostatných rizík, napríklad kreditné alebo tržné. Operačné riziko možno podľa Slovenského Zákona č. 644/2006 Z. z., ktorým sa mení a dopĺňa zákon č. 483/2001 definovať ako:

„Riziko vyplývajúce z nevhodných alebo chybných vnútorných postupov, zo zlyhania ľudského faktora, zo zlyhania používaných systémov alebo z vonkajších udalostí; súčasťou operačného rizika je právne riziko, ktoré znamená riziko vyplývajúce najmä z nevykonalnosti zmlúv, hrozby neúspešných súdnych konaní alebo rozsudkov s negatívnym vplyvom na banku“ [Zákon č. 644/2006 Z.z.].

Podľa Basel II sa operačné riziko definuje takto: „je to riziko straty vyplývajúce z nedostatočnosti alebo zlyhania vnútorných procesov, osôb, systémov alebo z vonkajších udalostí.“

Operačné riziko sa pritom nevzťahuje na produkciu alebo služby poskytované firmou (napríklad v banke operačné riziko nezahŕňa straty v dôsledku obchodovania, investovania, požičiavania peňazí a podobne), ale strata, ktorá vznikla jednaním pracovníka, ktorý pri obchodovaní prekročil povolený limit, bude zahrnutá pod straty, ktoré spadajú pod operačné riziko. Právne riziko vyskytujúce sa v definícii sa môže chápať ako riziko, keď protistrana nie je právne spôsobilá uzatvárať kontrakt, alebo riziko konfliktu transakcií v súlade s legislatívou. Pod operačné riziko avšak nespadá strategické riziko (riziko straty v dôsledku zlého obchodného rozhodnutia), ani reputačné riziko („straty vznikajúce pri poškodení reputácie, respektíve imidžu podnikateľského subjektu z dôvodu uskutočnených transakcií alebo zlyhaní podnikateľských procesov“) [Sivák, 2010].

Pojem operačné riziko sa môže aplikovať na akúkoľvek organizáciu alebo firmu, ale metódy pre meranie a modelovanie tohto rizika sa najviac rozvíjali v bankách (na základe dokumentu Basel II). V súčasnosti sa tento pojem začal používať a rozvíjať aj v poisťovníctve, a to hlavne na základe smernice Európskej únie pod názvom Solvency II.

Dá sa predpokladať, že v budúcnosti budú banky a poisťovne viac investovať do identifikácie, merania a riadenia operačného rizika. Veľká časť investícií pôjde pravdepodobne do systémov, rôznych softwarov a kontrol (hlavne interných), ktoré budú zamerané na zníženie operačných chýb a rizík.

2. Parametrické rozdelenia

2.1 LDA prístup

V rámci Basel II existujú tri metódy výpočtu kapitálových požiadaviek na krytie operačného rizika: Basic Indicator Approach (BIA), Standardized Approach (STA) a Advanced Measurement Approach (AMA). V rámci AMA existuje niekoľko vnútorných prístupov, z ktorých je pre nás najdôležitejší LDA (Loss Distribution Approach). Princípom tohto prístupu je konštrukcia distribučnej funkcie strát na základe historických informácií. Neočakávané straty z operačného rizika sa odvodzujú zo zloženého rozdelenia pravdepodobností.

Škodová udalosť X_i (používa sa označenie veľkosť škody) je udalosť, pri ktorej subjekt utrpí škodu, ktorú je možné vyjadriť v peňažnej hodnote. Celkovú stratu za nejaké obdobie je možné vyjadriť ako sumu

$$S = \sum_{i=1}^N X_i,$$

kde N je náhodná veličina, ktorá reprezentuje počet škôd za určité obdobie. Za predpokladu, že X_i sú nezávislé a rovnako rozdelené náhodné veličiny a každé X_i je nezávislé na N , potom má náhodná veličina S zložené rozdelenie. Ak označíme $P(x)$ distribučnú funkciu náhodných veličín X_i , $i = 1, 2, 3, \dots$, potom ako model pre $P(x)$ sa dá použiť známe rozdelenie pravdepodobností, ktorého parametre sa odhadnú na základe pozorovaných dát. Na modelovanie počtu škôd N sa používajú diskrétné rozdelenia, hlavne Poissonovo rozdelenie.

Riziko môže byť potom merané ako kvantil rozdelenia S . Basel II vyžaduje, aby to bol 99,9% kvantil [Dutta, Perry, 2007].

2.2 Parametrické rozdelenia pre veľkosti škôd

Jedným z najstarších prístupov na vytvorenie modelu rozdelenia strát je, aby parametrické rozdelenia čo najlepšie vystihovali dáta pre veľkosť a frekvenciu škôd. Parametre môžu byť odhadované pomocou rôznych techník ako napríklad metóda maximálnej vierohodnosti, metóda momentov, alebo pomocou odhadov kvantilov.

V tejto práci budú prezentované základné parametrické rozdelenia zamerané na veľkosť škôd, nie na frekvenciu škôd.

Spojité rozdelenia môžu mať rôzny počet parametrov. Rozdelenia, ktoré majú spravidla jeden až dva parametre je možné označiť ako jednoduché parametrické rozdelenia. Rozdelenia, ktoré majú tri alebo viac parametrov, sa môžu označiť ako zovšeobecnené parametrické rozdelenia a vhodnou voľbou parametrov sa môžu popísať špeciálne jednoduché rozdelenia.

Rozdelenie	Hustota $f(x)$	Počet parametrov
Exponenciálne	$f(x) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) I_{[0,\infty)}(x)$	Jeden
Weibullovo	$f(x) = \frac{\kappa}{\lambda} \left(\frac{x}{\lambda}\right)^{\kappa-1} \exp\left(-\left(\frac{x}{\lambda}\right)^{\kappa}\right) I_{[0,\infty)}(x)$	Dva
Gama	$f(x) = \frac{1}{\lambda^{\alpha} \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\lambda}\right) I_{[0,\infty)}(x)$	Dva
Logaritmicke – normálne	$f(x) = \frac{1}{x \sigma \sqrt{2\pi}} \exp\left[-\left(\frac{\ln x - \mu}{\sigma \sqrt{2}}\right)^2\right] I_{(0,\infty)}(x)$	Dva
Logaritmicke – Logistické	$f(x) = \frac{x^{\eta-1} \eta \alpha^{-\eta}}{\left[1 + \left(\frac{x}{\alpha}\right)^{\eta}\right]^2} I_{(0,\infty)}(x)$	Dva
Zovšeobecnené Pareto	$\frac{1}{\beta} \left(1 + \frac{\xi}{\beta} x\right)^{-\frac{1}{\xi}-1} I_{[0,\infty)}(x)$	Dva

Tabuľka 1: Vybrané jednoduché parametrické rozdelenia

Zdroj: Dutta K., Perry J.: *A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital*, 2007

Poznámka: $I_{[0,\infty)}(x) = 1$ pre $x \geq 0$ a 0 inak.

Definičným oborom hustoty väčšiny spojitých rozdelení sú buď všetky reálne čísla alebo všetky kladné reálne čísla. Niektoré rozdelenia ako napríklad beta rozdelenie, alebo rovnomerné rozdelenie majú obmedzený nosič. Pokiaľ sa modelujú straty,

ktoré nemôžu nadobudnúť záporné hodnoty, je treba uvažovať rozdelenia s kladným nosičom. Tabuľka 1 ukazuje vybrané parametrické rozdelenia, ktoré majú jeden alebo dva parametre.

Do úvahy je možné zobrať aj ďalšie rozdelenie, ktoré sa nazýva $g - h$ rozdelenie. Toto rozdelenie má 4 parametre. Viac o tomto rozdelení bude spomenuté v kapitole 3.

2.2.1 Exponenciálne rozdelenie

Ak má náhodná veličina X exponenciálne rozdelenie, potom hustota X je definovaná ako

$$f(x) = \frac{1}{\lambda} \exp\left(-\frac{x}{\lambda}\right) I_{[0, \infty)}(x), \quad (1)$$

kde parameter $\lambda > 0$. Exponenciálne rozdelenie je rozdelenie s jedným parametrom, ktoré sa často používa ako model pre náhodnú dobu čakania na nejakú udalosť (doba čakania na obsluhu, doba životnosti súčiastky, doba trvania nejakého deja) a to vtedy, ak rozdelenie zvyšnej doby čakania je nezávislé na tom, ako dlho čakáme; hovorí sa, že exponenciálne rozdelenie nemá pamäť, t.j.

$$P(X > s + t \mid X > t) = P(X > s) \text{ pre všetky } s, t \geq 0.$$

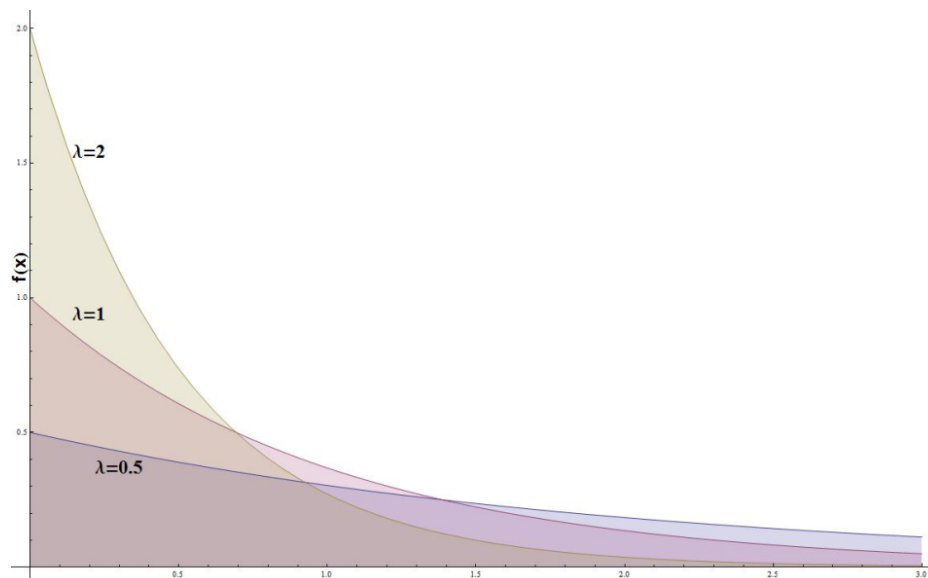
Exponenciálne rozdelenie je špeciálnym prípadom Gama rozdelenia ($\alpha = 1$) a Weibullovoho rozdelenia ($\kappa = 1$).

Základné popisné charakteristiky exponenciálneho rozdelenia:

Stredná hodnota: $E(X) = \lambda$

Rozptyl: $Var(X) = \lambda^2$

Značenie: $X \sim Exp(\lambda)$



Obr. 1: Hustota exponenciálneho rozdelenie pre hodnoty $\lambda = 0.5$; $\lambda = 1$; $\lambda = 2$

2.2.2 Weibullovo rozdelenie

Ak náhodná veličina má Weibullovo rozdelenie, potom hustota pravdepodobnosti má tvar :

$$f(x) = \frac{\kappa}{\lambda} \left(\frac{x}{\lambda}\right)^{\kappa-1} \exp - \left(\frac{x}{\lambda}\right)^{\kappa} I_{[0,\infty)}(x), \quad (2)$$

kde parametre $\kappa, \lambda \geq 0$. Toto rozdelenie má dva parametre a je flexibilnejšie, ako exponenciálne rozdelenie.

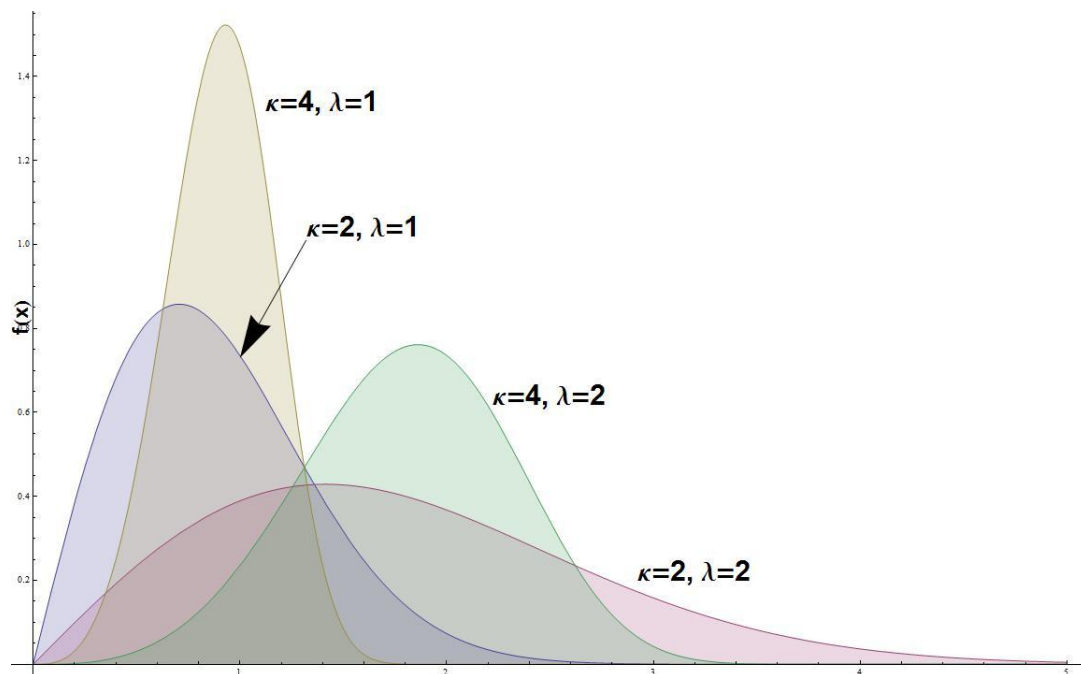
Weibullovo rozdelenie majú doby životnosti (doby do poruchy) strojných zariadení, u ktorých sa prejavuje mechanické opotrebovanie a únava materiálu (Klein, Bahýl, Vacek, [2002]). Používa sa na modelovanie poruchovosti, ktorá rastie ak $\kappa > 1$ (čím dlhšie stroj pracuje, tým je väčšia pravdepodobnosť, že behom nasledujúcich sekúnd sa pokazí), klesá ak $\kappa < 1$ (čím dlhšie stroj pracuje, tým je menšia pravdepodobnosť, že behom nasledujúcich sekúnd sa pokazí), alebo ostáva konštantná v čase, ak $\kappa = 1$ (nezáleží na tom, ako dlho stroj pracuje, a riziko, že sa stroj pokazí v nasledujúcich sekundách, ostáva rovnaké.). Ak $\kappa = 1$ dostávame exponenciálne rozdelenie.

Základné popisné charakteristiky Weibullovoho rozdelenia:

Stredná hodnota: $E(X) = \lambda \Gamma\left(1 + \frac{1}{\kappa}\right)$

Rozptyl: $Var(X) = \lambda^2 \left[\Gamma\left(1 + \frac{2}{\kappa}\right) - \Gamma^2\left(1 + \frac{1}{\kappa}\right) \right]$

Značenie: $X \sim W(\lambda, \kappa)$



Obr. 2: Hustota Weibullovo rozdelenia pre hodnoty $\kappa = 2; 4$

$\lambda = 1; 2$

2.2.3 Gama rozdelenie

Ak náhodná veličina X má Gama rozdelenie, potom hustota pravdepodobnosti má tvar:

$$f(x) = \frac{1}{\lambda^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\lambda}\right) I_{[0,\infty)}(x), \quad (3)$$

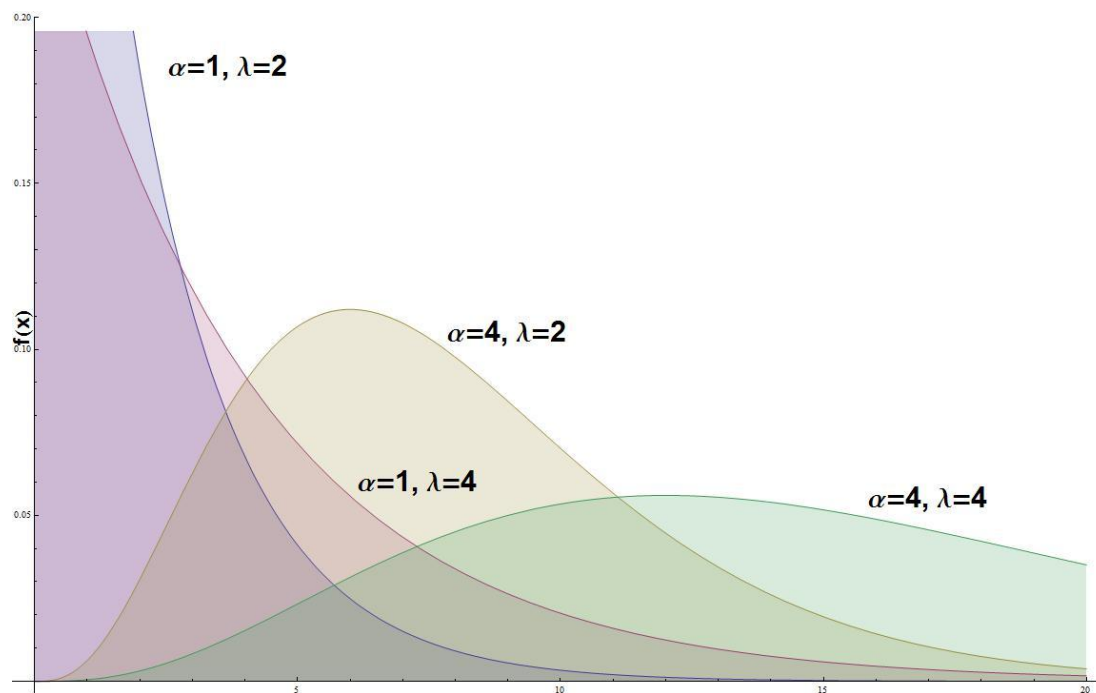
kde λ a α sú kladné a $\Gamma(\alpha)$ je Gama funkcia definovaná vzťahom: $\int_0^\infty t^{\alpha-1} e^{-t} dt$. Je možné ukázať, že ak $\{X_n\}$ je postupnosť nezávislých, exponenciálne rozdelených náhodných veličín s parametrom λ , potom $Y = \sum_{i=1}^n X_i$ má Gama rozdelenie s $\alpha = n$ a parametrom λ . Exponenciálne rozdelenie je špeciálnym prípadom Gama rozdelenia pre $\alpha = 1$ a rozdelenie chí – kvadrát o k stupňoch voľnosti je taktiež špeciálnym typom Gama rozdelenia pre $\alpha = 2k$ a $\lambda = 2$.

Základné popisné charakteristiky Gama rozdelenia:

Stredná hodnota: $E(X) = \alpha \lambda$

Rozptyl: $Var(X) = \alpha \lambda^2$

Značenie: $X \sim \Gamma(\alpha, \lambda)$ alebo $X \sim Gama(\alpha, \lambda)$



Obr. 3: Hustota Gama rozdelenia pre hodnoty $\alpha = 1; 4$
 $\lambda = 2; 4$

2.2.4 Logaritmicko – normálne rozdelenie

Ak Y je náhodná veličina s normálnym rozdelením, potom $X = e^Y$ má logaritmicko – normálne (Log - normálne) rozdelenie s hustotou pravdepodobnosti

$$f(x) = \frac{1}{x \sigma \sqrt{2\pi}} \exp \left[- \left(\frac{\ln x - \mu}{\sigma \sqrt{2}} \right)^2 \right] I_{(0,\infty)}(x), \quad (4)$$

kde $\mu \in \mathbb{R}$ je stredná hodnota a $\sigma > 0$ je štandardná odchýlka $Y = \ln X$. Dôležitou vlastnosťou logaritmicko - normálneho rozdelenia je, že ak $\{X_n\}$ je postupnosť nezávislých náhodných veličín s logaritmicko - normálnym rozdelením, kde pre každé X_i platí $X_i \sim LN(\mu, \sigma_i^2)$, potom

$$Y = \prod_{i=1}^n X_i \sim LN\left(\mu, \sum_{i=1}^n \sigma_i^2\right). \quad (5)$$

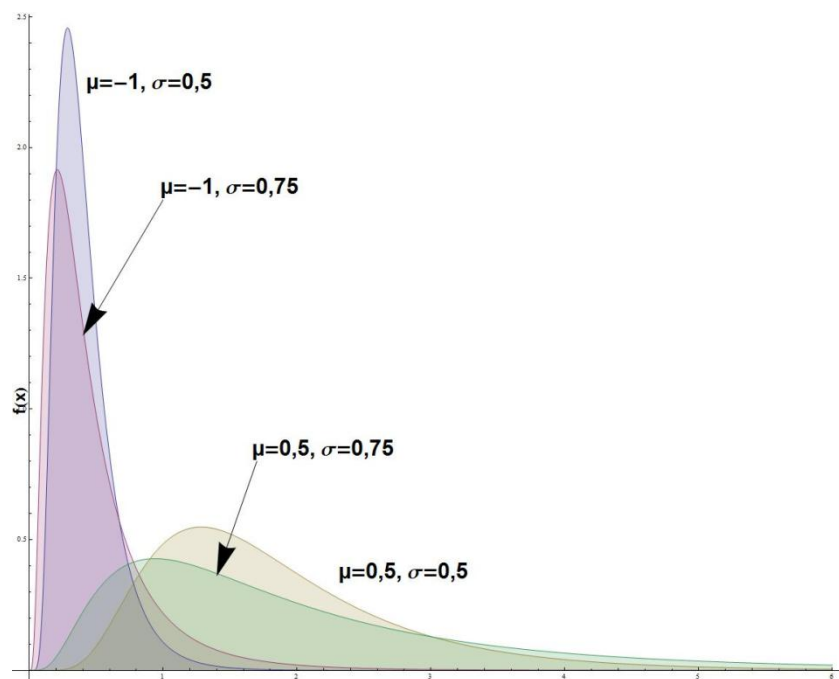
Táto vlastnosť je dôležitá, pretože ak určité straty majú logaritmicko – normálne rozdelenie, tak súčin týchto strát má taktiež logaritmicko – normálne rozdelenie.

Základné popisné charakteristiky logaritmicko – normálneho rozdelenia:

Stredná hodnota: $E(X) = \exp\left(\mu + \frac{\sigma^2}{2}\right)$

Rozptyl: $Var(X) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$

Značenie: $X \sim LN(\mu, \sigma)$



Obr. 4: Hustota Logaritmicko – Normálneho rozdelenia pre hodnoty $\mu = -1; 0,5$
 $\sigma = 0,5; 0,75$

2.2.5 Logaritmicko- logistické rozdelenie

Logaritmicko - logistické rozdelenie má nasledujúcu hustotu pravdepodobnosti pre náhodnú veličinu X:

$$f(x) = \frac{x^{\eta-1} \eta \alpha^{-\eta}}{\left[1 + \left(\frac{x}{\alpha}\right)^{\eta}\right]^2} I_{(0,\infty)}(x). \quad (6)$$

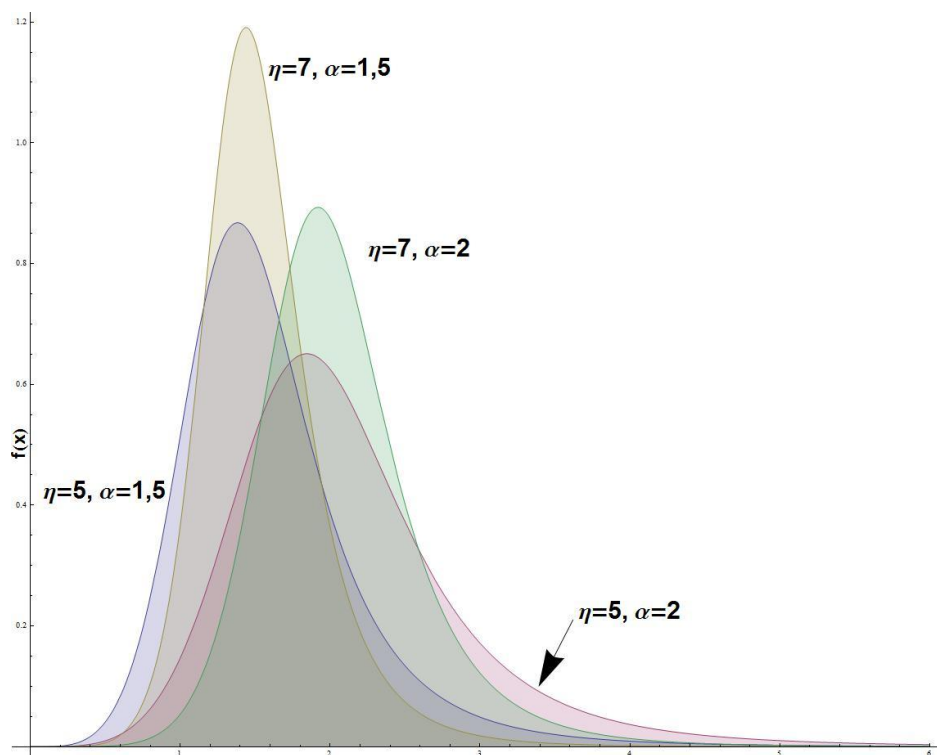
V tejto forme η slúži ako parameter tvaru a α je parameter polohy. Je možné pridať parameter škály, ale v tejto podobe je jeho hodnota jednotková. Náhodná veličina má logaritmicko – logistické rozdelenie, ak logaritmus z náhodnej veličiny je logistický rozdelený.

Základné popisné charakteristiky logaritmicko – logistického rozdelenia:

Stredná hodnota: $E(X) = \frac{\alpha \pi}{\eta \sin\left(\frac{\pi}{\eta}\right)}$

Rozptyl: $Var(X) = \alpha^2 \left(\frac{2 \pi}{\eta \sin\left(2 \frac{\pi}{\eta}\right)} - \frac{\pi^2}{\eta^2 \sin^2\left(\frac{\pi}{\eta}\right)} \right)$

Značenie: $X \sim LL(\eta, \alpha)$



Obr. 5: Hustota Logaritmicko – logistického rozdelenia pre hodnoty $\eta = 5; 7$

$$\alpha = 1,5; 2$$

2.2.6 Zovšeobecnené Paretovo rozdelenie

Rozdelenie s distribučnou funkciou $F(x)$

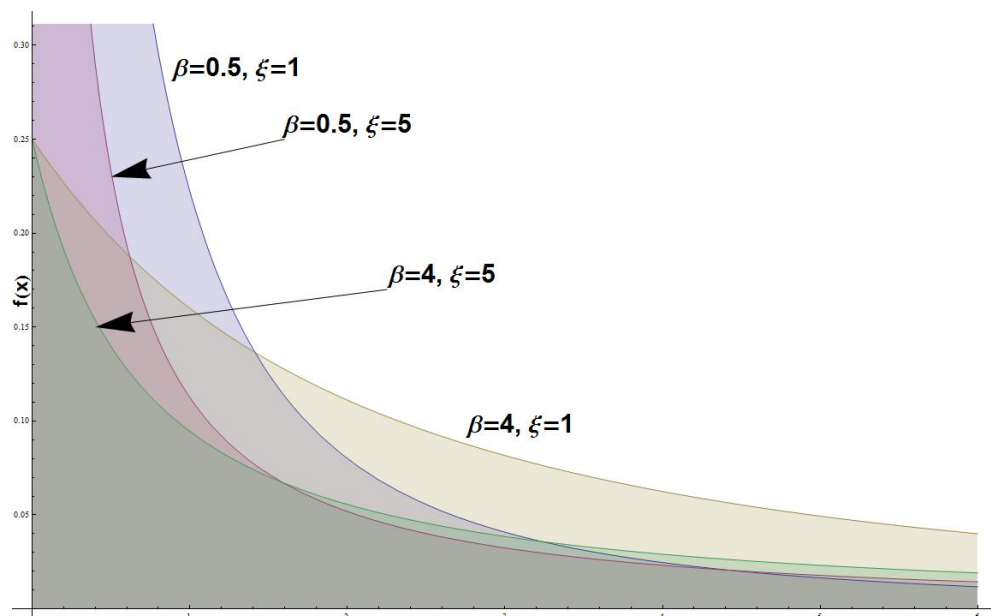
$$F(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}} & \text{keď } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{keď } \xi = 0 \end{cases} \quad (7)$$

kde $x \geq 0$ pre $\xi \geq 0$ a $x \in \left[0, -\frac{\beta}{\xi}\right]$ pre $\xi < 0$ bez parametra polohy, je nazývané ako zovšeobecnené Paretovo rozdelenie GPD (Generalized Pareto Distribution). Parameter $\beta > 0$ je parameter škály. Dôležitou vlastnosťou GPD je to, že ak veličina X má zovšeobecnené Paretovo rozdelenie pre $\xi > 0$, potom $E[X^k] = \infty$ pre $k \geq \frac{1}{\xi}$. Takže GPD nemá konečnú strednú hodnotu (a vyššie momenty) pre $\xi \geq 1$. Toto rozdelenie je veľmi dôležité pri modelovaní chvostov rozdelení v teórii extrémnych hodnôt EVT (Extreme Value Theory).

Stredná hodnota: $EX = \frac{\beta}{1-\xi}$, $\xi < 1$

Rozptyl: $Var(X) = \frac{\beta^2}{(1-\xi)^2(1-2\xi)}$, $\xi < 1/2$

Značenie: $X \sim GPD(\beta, \xi)$



Obr. 6: Hustota zovšeobecneného Paretovho rozdelenia pre hodnoty $\beta = 0,5; 4$

$\xi = 1; 5$

3. $G-h$ rozdelenie a jeho vlastnosti

$G-h$ rozdelenie prvýkrát predstavil americký štatistik John Tukey v roku 1977. Tomuto rozdeleniu sa venovalo viacero štatistikov. V tejto bakalárskej práci sa vychádza z dokumentov [Dutta, Perry, 2007], [Dutta, Babbel, 2002] a [Embrechts, Degen, Lambrigger, 2007].

3.1 Základná definícia

Tukey zaviedol množinu rozdelení transformovaním náhodnej veličiny Z , ktorá má normálne rozdelenie na intervale $(0,1)$ na

$$Y_{g,h}(Z) = (e^{gZ} - 1) \frac{\exp(hZ^2/2)}{g}, \quad (8)$$

kde g a h sú reálne parametre. V tejto práci sa za parametre g, h budú brať len kladné čísla, lebo inak transformácia (8) by už nebola rastúca pre určité hodnoty Z . Definičný obor $g-h$ rozdelenia zahŕňa všetky reálne čísla, kladné ako aj záporné. Preto veličina, ktorá má $g-h$ rozdelenie môže nadobúdať aj negatívne hodnoty a to aj vtedy, ak dáta použité pre toto rozdelenie sú pozitívne.

Po zavedení parametru polohy (A) a parametru škály (B) $g-h$ rozdelenie má 4 kladné parametre a má nasledujúcu formu

$$X_{g,h}(Z) = A + B (e^{gZ} - 1) \frac{\exp(hZ^2/2)}{g} = A + B \cdot Y_{g,h}(Z). \quad (9)$$

Značíme $X \sim g-h$ alebo, ak X má distribučnú funkciu F , tak $F \sim g-h$.

V prípade ak $h = 0$, $g-h$ rozdelenie (9) sa nám zredukuje na tvar

$$X_{g,0}(Z) = A + B \frac{(e^{gZ} - 1)}{g}, \quad (10)$$

ktorý je známy pod názvom $G-h$ rozdelenie. $G-h$ rozdelenie tak zodpovedá škálovému Logaritmickeo – normálnemu rozdeleniu. Parameter g je zodpovedný za šikmosť $G-h$ rozdelenia [Dutta, Perry, 2007].

Podobne keď $g = 0$, vzťah (9) limitným prechodom prejde na tvar

$$X_{0,h}(Z) = A + BZ \exp\left(\frac{hZ^2}{2}\right) = A + B \cdot Y_{0,h}(Z), \quad (11)$$

ktoré je známe ako H – rozdelenie. Parameter h v g – h rozdelení je zodpovedný za špicatost'. H - rozdelenie môže modelovať ťažké chvosty (špicatost'). Ak $g = h = 0$, $g - h$ rozdelenie konverguje k normálnemu rozdeleniu.

Keď $A = 0$, potom $X_{-g,h}(Z) = -X_{g,h}(-Z)$. To značí, že zmena znamienka g zmení iba smer, ale nie hodnotu šikmosti $g - h$ rozdelenia.

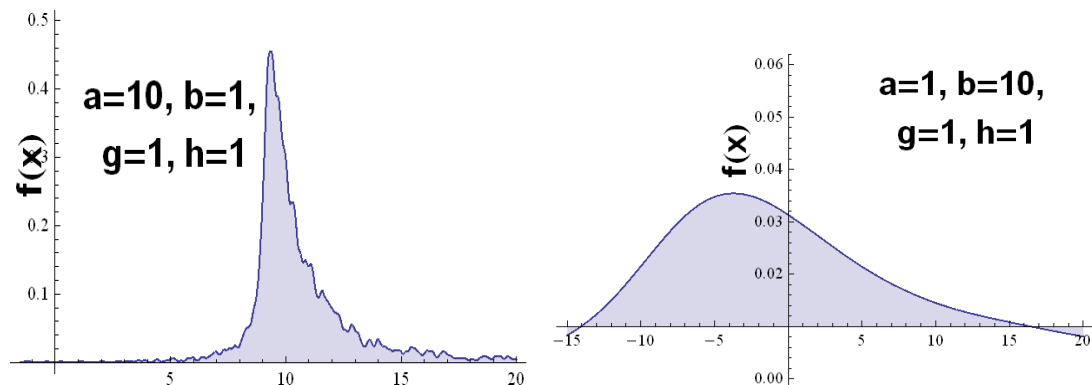
$G - h$ rozdelenie môže aproximovať širokú paletu rozdelení a to vybratím vhodných hodnôt za parametre A, B, g a h . Sila tohto rozdelenia je v schopnosti pravdepodobnostne aproximovať rozdelenia rôznych dát.

Poznámka: Pretože funkcia $k(x) = \frac{e^{gx}-1}{g} e^{hx^2/2}$ pre $h > 0$ je striktne rastúca, distribučná funkcia F náhodnej veličiny X , ktorá má $g - h$ rozdelenie je

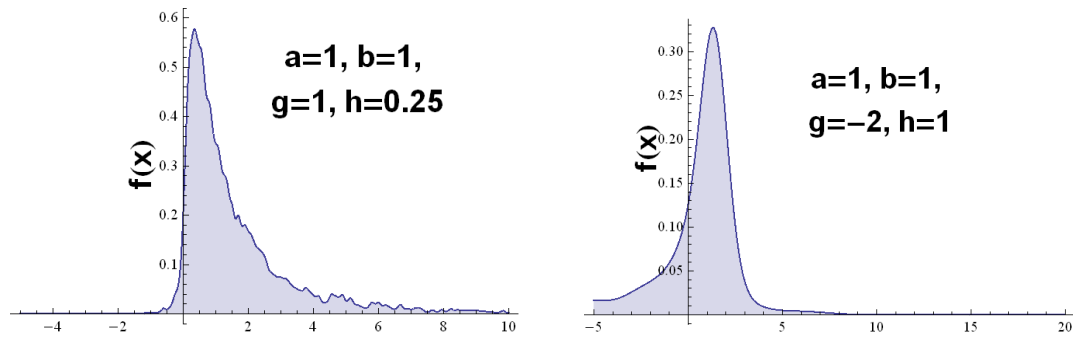
$$F(x) = \phi(k^{-1}(x)),$$

kde ϕ značí distribučnú funkciu normálneho rozdelenia. Hodnota v riziku (VaR) náhodnej veličiny X s $g - h$ rozdelením je

$$VaR_{\alpha}(X) = F^{-1}(\alpha) = k(\Phi^{-1}(\alpha)), 0 < \alpha < 1.^1$$



¹ VaR – Value at Risk



Obr. 7: Hustota $g - h$ rozdelenia

Na obrázku č. 7 sú zobrazené hustoty $g - h$ rozdelenia pre rôzne hodnoty parametrov a, b, g, h .

3.2 Odhad parametrov $g - h$ rozdelenia

$X_{g,h}(Z)$ je striktné rastúca transformácia náhodnej veličiny Z , ktorá má normálne rozdelenie na intervale $(0,1)$, so štyrmi parametrami $A, B, g, h \geq 0$.

Ak sa definuje X_p ako p -tý percentil $g - h$ rozdelenia a Z_p p -tý percentil normálneho rozdelenia. Potom

$$X_p - A = \frac{B}{g} (\exp(g Z_p) - 1). \quad (12)$$

Zmenením p na $1 - p$ dostaneme

$$X_{1-p} - A = \frac{B}{g} (\exp(g Z_{1-p}) - 1). \quad (13)$$

Platí, že $Z_p = -Z_{1-p}$ a použitím logaritmu na podiel (12) a (13) máme

$$g = -\left(\frac{1}{Z_p}\right) \ln \left(\frac{X_{1-p} - X_{0,5}}{X_{0,5} - X_p} \right). \quad (14)$$

Je treba poznamenať, že $Z_{0,5} = 0$. Nahradením vo vzorci (9) sa dostane $A = X_{0,5}$. Takto sa ukázala ďalšia vlastnosť $g - h$ rozdelenia, a to, že parameter polohy A je medián rozdelenia a parameter g je možné vyjadriť vzťahom (14). X_p , X_{1-p} a $X_{0,5}$ sú empirické kvantily. Táto vlastnosť implikuje to, že výberom rôznych hodnôt p je

možné získať rôzne odhady parametru g . Odhad tohto parametru teda môžeme napísať ako

$$\widehat{g}_p = -\left(\frac{1}{Z_p}\right) \ln\left(\frac{\widehat{X_{1-p}} - \widehat{X_{0,5}}}{\widehat{X_{0,5}} - \widehat{X_p}}\right)$$

Tieto odhady g pre rôzne hodnoty p odrážajú zmeny v šikmosti dát.

Pretože je veľa rôznych odhadov g_p ktoré závisia na percentile p , v tejto práci sa za g berie medián, tak ako je to odporúčané v práci [Dutta, Perry, 2007].

Použitím rovnosti (9) sa dostane vyjadrenie percentilu $g - h$ rozdelenia

$$X_p = A + B(e^{g Z_p} - 1) \frac{\exp(h Z_p^2/2)}{g}, \quad (15)$$

$$X_{1-p} = A + B(e^{g Z_{1-p}} - 1) \frac{\exp(h Z_{1-p}^2/2)}{g}. \quad (16)$$

Aplikovaním $Z_p = -Z_{1-p}$ a odčítaním (16) od (15) sa dostane:

$$\frac{g(X_p - X_{1-p})}{(e^{g Z_p} - e^{-g Z_p})} = B \exp(h Z_p^2/2), \quad (17)$$

respektíve

$$\ln \frac{g(X_p - X_{1-p})}{(e^{g Z_p} - e^{-g Z_p})} = \ln(B) + h (Z_p^2/2). \quad (18)$$

Takže takto možno získať hodnotu h z lineárnej regresie $\ln \frac{g(X_p - X_{1-p})}{(e^{g Z_p} - e^{-g Z_p})}$ na $Z_p^2/2$.

$e^{\ln(B)}$, kde $\ln(B)$ je odhad získaný regresiou, je odhadom B a koeficient regresie dá odhad pre hodnotu h . Tak sa získajú hodnoty h a B , ktoré sú podmienené na hodnote g . Ak sa použijú oddelene dve polovice dát, možno získať dva rozdielne odhady parametru h a tak dva rôzne odhady špicatosti pre dva chvosty rozdelenia. Tento postup sa môže opakovať, ak v (15) respektíve v (16) zvolíme $p = 1/2$, tak dostaneme vzťah pre Upper half spread (UHS) respektíve pre Lower half spread (LHS), ktoré sú definované nasledovne:

$$UHS = \frac{g(X_{1-p} - X_{0,5})}{(e^{-g Z_p} - 1)} \quad (19)$$

$$LHS = \frac{g(X_{0,5} - X_p)}{(1 - e^{g Z_p})}. \quad (20)$$

Nahradením ľavej strany v rovnici (17) výrazmi UHS, respektíve LHS, sa dostane odhad h založený na hornú, respektíve dolnú polovicu dát.

Poznámka: Niekedy sa stáva, že transformácia vyjadrená v (9) nedovolí $g - h$ rozdeleniu popisovať dáta veľmi dobre. V takýchto prípadoch parametre g a h môžu byť vyjadrené ako polynomicke funkcie Z^2 . Všeobecnejšia forma $g - h$ transformácie je

$$X_{g,h}(Z) = A + B (e^{g(Z^2)Z} - 1) \frac{\exp(h(Z^2)Z^2/2)}{g(Z)}, \quad (21)$$

kde $g(Z^2)$ a $h(Z^2)$ sú polynómy Z^2 . Napríklad: $g(Z^2) = \gamma_0 + \gamma_1 Z^2 + \gamma_2 Z^4 + \gamma_3 Z^6$ a $h(Z^2) = \eta_0 + \eta_1 Z^2 + \eta_2 Z^4 + \eta_3 Z^6$.

3.3 Momenty $g - h$ rozdelenia

Z rovnice (9) platí $X^n = (A + BY)^n = \sum_{i=0}^n \binom{n}{i} A^{n-i} B^i Y^i$, čiže

$$E(X^n) = \sum_{i=0}^n \binom{n}{i} A^{n-i} B^i E(Y^i), \quad (22)$$

kde $Y = (e^{gZ} - 1) \frac{\exp(hZ^2/2)}{g}$.

Pre $g \neq 0$ je $Y^i = \frac{\exp(i h Z^2/2) \sum_{r=0}^i \binom{i}{r} (-1)^r \exp(-(i-r)gZ)}{g^i}$, kde $Z \sim N(0,1)$. Takže pre strednú hodnotu Y platí nasledujúci vzťah

$$E(Y^i) = \frac{\sum_{r=0}^i \binom{i}{r} (-1)^r \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(1-ih)z^2 + (i-r)gz\right) dz}{\sqrt{2\pi} g^i}. \quad (23)$$

Integrál vo výraze (23) sa dá upraviť na tvar

$$\begin{aligned} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(1-ih)z^2 + (i-r)gz\right) dz \\ = \int_{-\infty}^{\infty} \exp\left(-\frac{((1-ih)z - (i-r)g)^2}{2(1-ih)}\right) \exp\left(\frac{(i-r)^2 g^2}{2(1-ih)}\right) dz \\ = \frac{\sqrt{2\pi}}{\sqrt{1-ih}} \exp\left(\frac{(i-r)^2 g^2}{2(1-ih)}\right). \end{aligned}$$

Aplikovaním tejto úpravy sa výraz (23) dá napísať ako

$$E(Y^i) = \frac{\sum_{r=0}^i (-1)^r \binom{i}{r} \exp\left(\frac{[(i-r)g]^2}{2(1-ih)}\right)}{g^i \sqrt{1-ih}}.$$

Dosadením do (22) dostaneme vzťah pre n – tý moment $g - h$ rozdelenia

$$E(X^n) = \sum_{i=0}^n \binom{n}{i} A^{n-i} B^i \frac{\sum_{r=0}^i (-1)^r \binom{i}{r} \exp\left(\frac{[(i-r)g]^2}{2(1-ih)}\right)}{g^i \sqrt{1-ih}}. \quad (24)$$

Tento vzťah je definovaný pre $0 \leq h \leq \frac{1}{n}$. Pre vyjadrenia strednej hodnoty, rozptylu, šikmosti a špicatosti sa pre jednoduchosť bude brať $A = 0$ a $B = 1$ [Dutta, 2007].

3.3.1 Stredná hodnota

Stredná hodnota $g - h$ rozdelenia pre $0 \leq h \leq 1$ sa zo vzťahu (24) rovná

$$E(X) = \mu = \frac{\exp\left(\frac{g^2}{2(1-h)}\right) - 1}{g\sqrt{1-h}}.$$

3.3.2 Rozptyl

Vzťah pre rozptyl náhodnej veličiny X je $\text{var}(X) = E[(X - EX)^2] = EX^2 - (EX)^2$.

$$E(X^2) = \frac{1}{g^2 \sqrt{1-2h}} \left[\exp\left(\frac{2g^2}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right) + 1 \right].$$

Rozptyl $g - h$ rozdelenie pre $0 \leq h \leq \frac{1}{2}$ je rovný

$$var(X) = \frac{1}{g^2\sqrt{1-2h}} \left[\exp\left(\frac{2g^2}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right) + 1 \right] - \left[\frac{\exp\left(\frac{g^2}{2(1-h)}\right) - 1}{g\sqrt{1-h}} \right]^2.$$

3.3.3 Šikmost'

Šikmost' náhodnej veličiny X je $\alpha_3 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{\mu_3^2}{\mu_2^3}$, kde $\mu_2 = var(X)$ a $\mu_3 = E[(X - EX)^3] = 2(EX)^3 - 3EX EX^2 + EX^3$.

$$E(X^3) = \frac{\exp\left(\frac{9g^2}{2(1-3h)}\right) - 3\exp\left(\frac{4g^2}{2(1-3h)}\right) + 3\exp\left(\frac{g^2}{2(1-3h)}\right) - 1}{g^3\sqrt{1-3h}}.$$

Šikmost' $g - h$ rozdelenia pre $0 \leq h \leq \frac{1}{3}$ po úpravách a zjednodušení je

$$\alpha_3 = \left[2 \left(\frac{\exp\left(\frac{g^2}{2(1-h)}\right) - 1}{\sqrt{1-h}} \right)^3 - 3 \left(\frac{\exp\left(\frac{g^2}{2(1-h)}\right) - 1}{\sqrt{1-h}} \right) \left(\frac{\exp\left(\frac{2g^2}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right) + 1}{\sqrt{1-2h}} \right) + \left(\frac{\exp\left(\frac{9g^2}{2(1-3h)}\right) - 3\exp\left(\frac{4g^2}{2(1-3h)}\right) + 3\exp\left(\frac{g^2}{2(1-3h)}\right) - 1}{\sqrt{1-3h}} \right) \right]^2 / \left[\frac{1}{\sqrt{1-2h}} \left[\exp\left(\frac{2g^2}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right) + 1 \right] - \left[\frac{\exp\left(\frac{g^2}{2(1-h)}\right) - 1}{\sqrt{1-h}} \right]^2 \right]^3.$$

3.3.4 Špicatost'

Špicatost' náhodnej veličiny X je $\alpha_4 = \frac{\mu_4}{\mu_2^2}$, kde $\mu_4 = E[(X - EX)^4] = -3(EX)^4 + 6(EX)^2 EX^2 - 4EX EX^3 + EX^4$.

$$E(X^4) = \frac{\exp\left(\frac{16g^2}{2(1-4h)}\right) - 4\exp\left(\frac{9g^2}{2(1-4h)}\right) + 6\exp\left(\frac{4g^2}{2(1-4h)}\right) - 4\exp\left(\frac{g^2}{2(1-4h)}\right) + 1}{g^4\sqrt{1-4h}}.$$

Špicatost' $g - h$ rozdelenia pre $0 \leq h \leq \frac{1}{4}$ je po úpravách a zjednodušení rovná

$$\begin{aligned} \alpha_4 &= \left[-3 \left(\frac{\exp\left(\frac{g^2}{2(1-h)}\right) - 1}{\sqrt{1-h}} \right)^4 \right. \\ &\quad + 6 \left(\frac{\exp\left(\frac{g^2}{2(1-h)}\right) - 1}{\sqrt{1-h}} \right)^2 \left(\frac{\exp\left(\frac{2g^2}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right) + 1}{\sqrt{1-2h}} \right) \\ &\quad - 4 \left(\frac{\exp\left(\frac{g^2}{2(1-h)}\right) - 1}{\sqrt{1-h}} \right) \left(\frac{\exp\left(\frac{9g^2}{2(1-3h)}\right) - 3\exp\left(\frac{4g^2}{2(1-3h)}\right) + 3\exp\left(\frac{g^2}{2(1-3h)}\right) - 1}{\sqrt{1-3h}} \right) \\ &\quad \left. + \left(\frac{\exp\left(\frac{16g^2}{2(1-4h)}\right) - 4\exp\left(\frac{9g^2}{2(1-4h)}\right) + 6\exp\left(\frac{4g^2}{2(1-4h)}\right) - 4\exp\left(\frac{g^2}{2(1-4h)}\right) + 1}{\sqrt{1-4h}} \right) \right] \\ &\quad / \left[\left(\frac{\exp\left(\frac{2g^2}{1-2h}\right) - 2\exp\left(\frac{g^2}{2(1-2h)}\right) + 1}{\sqrt{1-2h}} \right) - \left(\frac{\exp\left(\frac{g^2}{2(1-h)}\right) - 1}{\sqrt{1-h}} \right)^2 \right]^2. \end{aligned}$$

3.4 Chovanie chvostov

Štatistické vlastnosti pre odhady vysokých kvantilov vo veľkej miere závisia na správaní sa chvostu základného modelu. $G - h$ rozdelenie je v tomto ohľade veľmi flexibilné. Popisovaním vlastností ťažkých chvostov štatistických modelov sa zaoberal srbský matematik Jovan Karamata. V tomto texte sa bude vychádzať z dokumentu [Embrechts, 2007].

Definícia 1. Merateľná funkcia $L: R \rightarrow (0, \infty)$ sa nazýva pomaly sa meniaci v nekonečnu (SV - Slowly varying), ak pre $t > 0$ platí:

$$\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1.$$

Značí sa $L \in SV$. Funkcia f sa nazýva pravidelne sa meniaci (RV - Regularly varying) v nekonečnu s indexom $\alpha \in \mathbb{R}$, ak $f(x) = x^\alpha L(x)$. Značí sa $f \in RV_\alpha$. Platí $RV_0 = SV$.

Tvrdenie 1. Nech F je absolútne spojitá distribučná funkcia s hustotou f . Označme $\bar{F}(x) = 1 - F(x)$. Ak

$$\lim_{x \rightarrow \infty} \frac{x f(x)}{\bar{F}(x)} = \alpha > 0,$$

potom $\bar{F} \in RV_{-\alpha}$.

Toto tvrdenie poskytuje jednoduchý nástroj pre kontrolu, či daná funkcia je alebo nie je pravidelne sa meniaci v nekonečnu.

Dá sa ukázať, že $g - h$ rozdelenie je RV v nekonečnu s indexom $-1/h$ (stále sa predpokladá, že $h > 0$). Predpokladajme, že $X \sim g - h$, potom

$$F(x) = \phi(k^{-1}(x)), \quad f(x) = \frac{\varphi(k^{-1}(x))}{k'(k^{-1}(x))},$$

kde φ značí hustotu štandardného normálneho rozdelenia a ϕ distribučnú funkciu štandardného normálneho rozdelenia. Pre $g - h$ rozdelenie máme

$$\lim_{x \rightarrow \infty} \frac{x f(x)}{\bar{F}(x)} = \lim_{x \rightarrow \infty} \frac{x \varphi(k^{-1}(x))}{(1 - \phi(k^{-1}(x))) k'(k^{-1}(x))}. \quad (25)$$

Použitím substitúcie $u = k^{-1}(x)$ a $k(u) = \frac{e^{gu} - 1}{g} e^{hu^2/2}$ predchádzajúci výraz (25)

možno upraviť na tvar

$$\lim_{x \rightarrow \infty} \frac{x \varphi(k^{-1}(x))}{(1 - \phi(k^{-1}(x))) k'(k^{-1}(x))} = \lim_{u \rightarrow \infty} \frac{k(u) \varphi(u)}{(1 - \phi(u)) k'(u)} =$$

$$\begin{aligned}
&= \lim_{u \rightarrow \infty} \frac{\varphi(u) \frac{(e^{gu} - 1) \exp\left(\frac{hu^2}{2}\right)}{g}}{\exp\left(\frac{hu^2}{2}\right) \left(e^{gu} + \frac{(e^{gu} - 1)}{g} hu\right) (1 - \phi(u))} \\
&= \lim_{u \rightarrow \infty} \frac{\varphi(u)(e^{gu} - 1)}{(1 - \phi(u))(ge^{gu} + hu(e^{gu} - 1))}.
\end{aligned}$$

Použitím vzťahu $u(1 - \phi(u))/\varphi(u) \rightarrow 1$, pre $u \rightarrow \infty$ platí

$$\lim_{x \rightarrow \infty} \frac{x f(x)}{\bar{F}(x)} = \frac{1}{h},$$

a teda podľa Tvrdenia 1 $\bar{F} \in RV_{-1/h}$.

Takže ak $X \sim g - h$ pre $h > 0$, potom platí $P(X > x) = \bar{F}(x) = x^{-\frac{1}{h}} L(x)$ pre nejakú funkciu $L \in SV$. Správanie sa chvostov je teda dané chovaním funkcie $L(x)$ pre $x \rightarrow \infty$. Dá sa odvodiť explicitný asymptotický vzorec pre túto funkciu L pre $g - h$ rozdelenie. Jej tvar je nasledovný:

$$L(x) = \frac{1}{\sqrt{2\pi}g^{1/h}} \frac{\left[\exp\left(g\left(-\frac{g}{h} + \frac{1}{h}\sqrt{g^2 + 2h \log(gx)}\right)\right) - 1 \right]^{1/h}}{-\frac{g}{h} + \frac{1}{h}\sqrt{g^2 + 2h \log(gx)}} \left(1 + o\left(\frac{1}{\log x}\right)\right).$$

Vzhľadom k náročnosti odvodu tohto tvaru, nie je v tejto práci uvedené, ale je možné ho nájsť v dokumente [Embrechts, 2007].

3.4.1 Subexponenciálne rozdelenia

Definícia 2. Nech X_1, \dots, X_n sú nezávislé rovnako rozdelené náhodné veličiny s distribučnou funkciou F . Potom distribučná funkcia súčtu $X_1 + X_2 + \dots + X_n$ sa nazýva n - násobná konvolúcia distribučnej funkcie F .

Značenie: $F^{n*}(x) = P(X_1 + X_2 + \dots + X_n \leq x)$.

Definícia 3. Nech X_1, \dots, X_n sú nezávislé, nezáporné rovnako rozdelené náhodné veličiny s distribučnou funkciou F . Potom rozdelenie s distribučnou funkciou F je subexponenciálne, ak pre každé $n \geq 2$ platí

$$\lim_{x \rightarrow \infty} \frac{\overline{F^{n*}}(x)}{\overline{F}(x)} = n. \quad (26)$$

Zo vzťahu (26) sa dá odvodiť, že pre súčet $S_n = \sum_{i=1}^n X_i$, n nezávislých, nezáporných, rovnako rozdelených náhodných veličín X_1, \dots, X_n so subexponenciálnym rozdelením platí

$$\lim_{x \rightarrow \infty} \frac{P(S_n > x)}{P(\max_{1 \leq i \leq n} X_i > x)} = 1. \quad (27)$$

Čitateľ vo výraze (27) je rovný $1 - F^{n*}(x)$, čo sa rovná $\overline{F^{n*}}(x)$. Menovateľ vo výraze (27) sa dá ďalej upraviť:

$$\begin{aligned} P\left(\max_{1 \leq i \leq n} X_i > x\right) &= 1 - P\left(\max_{1 \leq i \leq n} X_i < x\right) = 1 - P(X_1 < x, X_2 < x, \dots, X_n < x) \\ &= 1 - \prod_{i=1}^n P(X_i < x) = 1 - F^n(x) \\ &= (1 - F(x))(1 + F(x) + \dots + F^{n-1}(x)). \end{aligned}$$

Takže

$$\lim_{x \rightarrow \infty} \frac{P(S_n > x)}{P(\max_{1 \leq i \leq n} X_i > x)} = \lim_{x \rightarrow \infty} \frac{\overline{F^{n*}}(x)}{\overline{F}(x)(1 + F(x) + \dots + F^{n-1}(x))} = 1.$$

Vzťah (27) vyjadruje skutočnosť, že chvost rozdelenia celkovej straty S_n je určený chvostom rozdelenia maximálnej straty.

4. Teória extrémnych hodnôt

Teória extrémnych hodnôt je nástroj, ktorý je vhodný na modeláciu extrémnych zriedkavých udalostí, ktoré môžu mať veľký dopad. Zaoberá sa hlavne koncami rozdelení, chvostmi.

4.1 POT metóda

Existujú dva základné spôsoby analýzy dát založené na EVT. Prvým je metóda blokového maxima, ktorej sa v tejto práci nebudeme venovať. Druhým a pre nás zaujímavejším spôsobom je metóda POT (z angličtiny Peaks over threshold). Táto metóda slúži k odhadovaniu pravdepodobností prekročenia veľmi vysokých hraníc (prahov), alebo k odhadovaniu vysokých kvantilov. Dáta nad hranicou indikujú extrémnu udalosť. Dôležitú úlohu tu zohráva zovšeobecnené Paretovo rozdelenie, pretože ak sa vhodne zvolí daná hranica, rozdelenie excesov sa dá aproximovať týmto rozdelením [Sanders, 2002].

Nech X_1, \dots, X_n sú nezávislé rovnako rozdelené náhodné veličiny, u nech je hranica (prah). Vyberieme tie hodnoty, ktoré prekračujú danú hranicu, ich počet si označíme ako N_u . Z týchto hodnôt určíme veľkosti excesov $Y_j = X_j - u$.

Distribučná funkcia zovšeobecneného Paretovho rozdelenia s nulovým parametrom polohy je

$$F_{\xi, \beta}(x) = 1 - \left(1 + \frac{\xi}{\beta}x\right)^{-\frac{1}{\xi}}, \xi \neq 0, \quad (28)$$

kde $x \geq 0$ pre $\xi \geq 0$ a $x \in \left[0, -\frac{\beta}{\xi}\right]$ pre $\xi < 0$.

Distribučná funkcia F_u rozdelenia hodnôt, ktoré v pozorovaní veličín s distribučnou funkciou F presahujú stanovenú hranicu (prah) u sa definuje ako

$$F^{[u]}(x) = P(X \leq x | X > u) = \frac{F(x) - F(u)}{1 - F(u)}, x \geq u. \quad (29)$$

Distribučná funkcia rozdelenia excesov, ktoré prekročia hranicu u je

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(x + u) - F(u)}{1 - F(u)}, 0 \leq x \leq x_F - u, \quad (30)$$

kde $x_F \leq \infty$ vyjadruje pravý koniec F . Podľa Pickands – Balkema – de Haanovej vety, distribučná funkcia (30) pre vhodnú hranicu u je dobre aproximovateľná distribučnou funkciou (28) zovšeobecneného Paretovho rozdelenia [Dutta, Perry, 2007].

Strednú hodnotu veľkosti prekročenia hranice u za podmienky, že dôjde k tomuto prekročeniu je možné definovať ako

$$e(u) = E(X - u | X > u).$$

Je to stredná hodnota rozdelenia s distribučnou funkciou (30). Pre náhodnú veličinu so zovšeobeným Paretovým rozdelením (28) majú excesy rozdelenie s distribučnou funkciou (30), s rovnakým parametrom ξ a s parametrom β závislým na hranici u [Mandl, Mazurová, Justová, 2010]. Platí teda:

$$F_u(x) = F_{\xi, \beta(u)}(x), \text{ pričom } \beta(u) = \beta + \xi u \quad (31)$$

a $x \geq 0$ pre $\xi \geq 0$ a $x \in \left[0, -\frac{\beta}{\xi} - u\right]$ pre $\xi < 0$.

Stredná hodnota zovšeobeného Paretovho rozdelenia

$$EX = \frac{\beta}{1 - \xi}, \quad (32)$$

je konečná pre $\xi < 1$. Zo vzťahov (31) a (32) platí

$$e(u) = \frac{\beta + \xi u}{1 - \xi},$$

kde $u \geq 0$ pre $0 \leq \xi < 1$ a $u \in \left[0, -\frac{\beta}{\xi}\right]$ pre $\xi < 0$.

Z n pozorovaní sa dá pre danú hranicu u zostrojiť empirický odhad strednej hodnoty

$$e_n(u) = \frac{\sum_{i=1}^n (X_i - u) I_{[X_i > u]}}{\sum_{i=1}^n I_{[X_i > u]}},$$

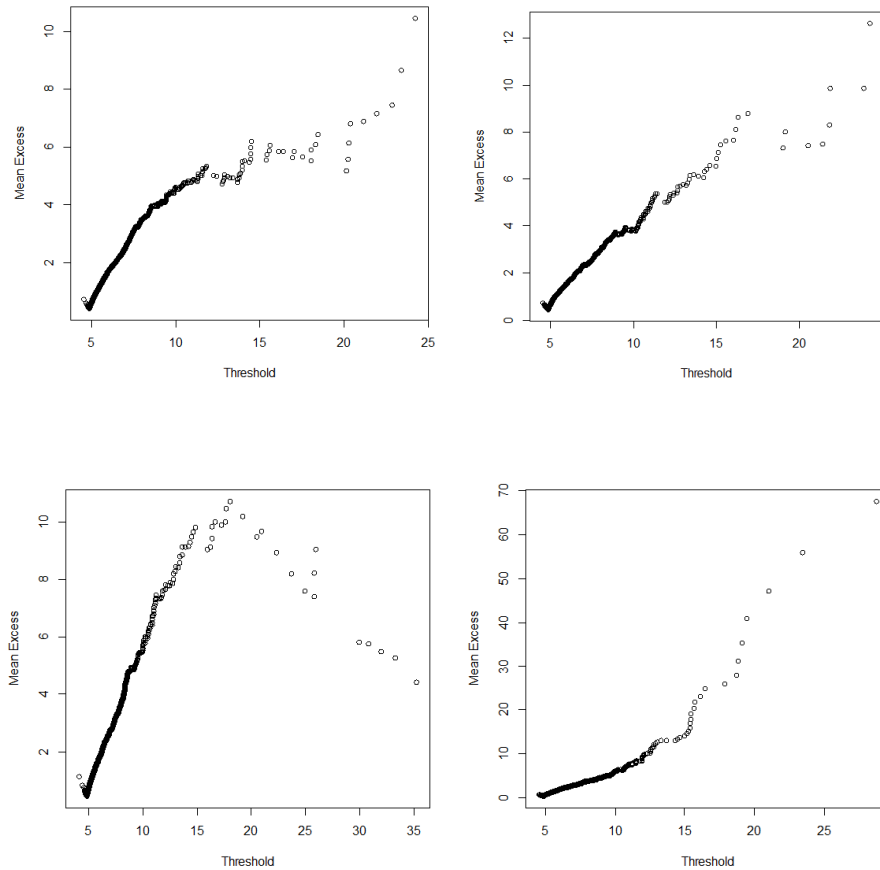
čo je vlastne suma excesov nad hranicou u delená počtom dát, ktoré prekračujú túto hranicu.

Dôležitou úlohou v POT metóde je zvolenie správnej hranice (prahu). Jednou z možností a veľmi užitočným nástrojom je mean excess plot (meplot):

$$\{(X_{i,n}, e_n(X_{i,n})), i = 2, \dots, n\},$$

kde $X_{n,n} \leq \dots \leq X_{i,n}$ sú poradové štatistiky z pozorovaní. Ak tieto pozorovania pochádzajú z GPD, tak by mal byť priebeh tohto grafu lineárny. Ak priebeh grafu z empirických hodnôt sa zdá byť od nejakého miesta u približne lineárny, tak to značí, že excesy prekračujúce danú hranicu u možno aproximovať zovšeobecneným Paretovým rozdelením s pozitívnym parametrom β .

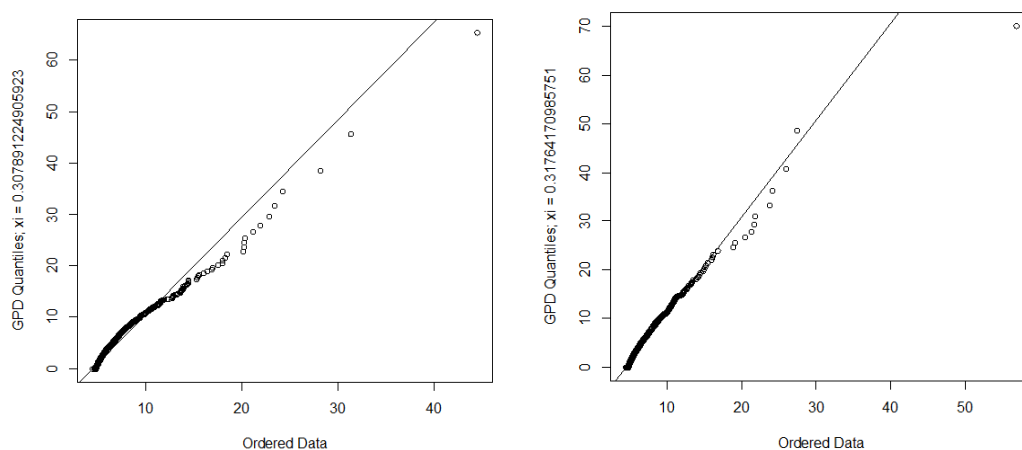
Na obrázku 8 sú nakreslené meploty pre $g - h$ rozdelenie s parametrami $a = 5$; $b = 0,2$; $g = 1,5$; $h = 0,2$. Počet nasimulovaných dát je 10 000. Okrem pár extrémnych bodov, ktoré sú priemerov len zopár veľkých excesov, priebehy grafov sa zdajú byť celkom lineárne, čo značí, že GPD bude dobre aproximovať tieto excesy.



Obr. 8: Grafy stredných hodnôt excesov $g - h$ rozdelenia pre parametre

$$a = 5; b = 0,2; g = 1,5; h = 0,2.$$

Na overenie, či daná hranica je vhodná posluží takzvaný QQplot (kvantil – kvantilový graf). Ide v podstate o graf, ktorý porovnáva dve rozdelenia pravdepodobností tým, že zobrazí ich kvantily (proti sebe). V našom prípade kvantily GPD rozdelenia s kvantilmi empirických hodnôt. Výborná zhoda by bola na priamke $y = x$. Ak je graf QQplotu blízko tejto priamke, tak to značí dobrú zhodu. Na obrázku 9 sú zobrazené QQploty pre prvé dva grafy z obrázka 8. V prvom prípade bola zvolená hranica 8, v tom druhom bola hranica na úrovni 8,5. Je vidieť, že vysoké kvantily ležia pod osou $y = x$ a to značí, že toto rozdelenie má ťažký chvost [Sanders, 2002].



Obr. 9: QQPlot

4.2 Odhady kvantilov

Cieľom praktickej časti tejto bakalárskej práce je určenie vysokých kvantilov $g - h$ rozdelenia, ako napríklad $x_{0,999}$. Ten reprezentuje hodnotu prekročenia v priemere jednu škodu z 1000.

Pre distribučnú funkciu F si definujeme kvantilovú funkciu F^{-1} ako

$$F^{-1}(p) = \inf\{x \in R: F(x) \geq p\}, \text{ kde } 0 < p < 1.$$

Ak rozdelenie poznáme, hodnotu $F^{-1}(p)$ vieme vypočítať. Avšak ak rozdelenie nepoznáme, tak túto hodnotu by sme chceli odhadnúť. Ak pre distribučnú funkciu F

budeme aproximovať rozdelenia excesov GPD rozdelením (28), dá sa pre $x \geq u$ písať

$$\bar{F}(x) = \bar{F}(u)\bar{F}^{(u)}(x-u) = \bar{F}(u)\left(1 + \frac{\xi}{\beta}(x-u)\right)^{-\frac{1}{\xi}}.$$

Inverziou sa dostane vyjadrenie kvantilu (označovaného niekedy ako Var_α) pre $\alpha \geq F(u)$

$$q_\alpha(F) = F^{-1}(\alpha) = u + \frac{\beta}{\xi} \left(\left(\frac{1-\alpha}{\bar{F}(u)} \right)^{-\xi} - 1 \right).$$

Pre $x \geq u$ je odhad

$$\hat{\bar{F}}(x) = \frac{N_u}{n} \left(1 + \frac{\hat{\xi}}{\hat{\beta}}(x-u) \right)^{-\frac{1}{\hat{\xi}}},$$

Kde N_u je počet pozorovaní prekračujúcich hranicu u a n je celkový počet pozorovaní. Takže pre $\alpha > 1 - N_u/n$ je odhad

$$\hat{q}_\alpha(F) = u + \frac{\hat{\beta}}{\hat{\xi}} \left(\left(\frac{1-\alpha}{N_u/n} \right)^{-\hat{\xi}} - 1 \right). \quad (33)$$

$\bar{F}(u)$ sa odhaduje pomerom N_u/n a parametre $\hat{\beta}$ a $\hat{\xi}$ sú odhady parametrov GPD.

Takto sme získali vzťah pre odhad kvantilov.

4.2.1 Aplikácia POT metódy

V tomto príklade porovnáme hodnoty kvantilov odhadnutých pomocou metódy POT s kvantilmi $g-h$ rozdelenia.

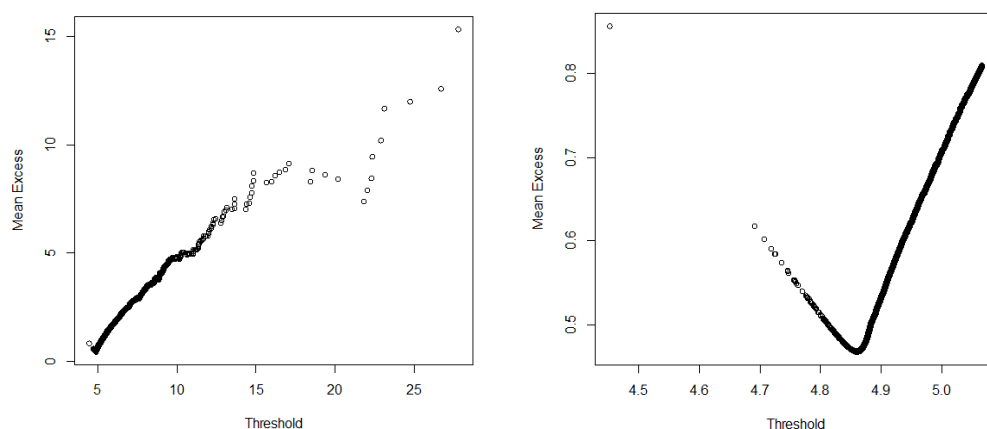
Nasimulujeme 10 000 hodnôt z $g-h$ rozdelenia s parametrami $a = 5; b = 0,2; g = 1,5; h = 0,2$. Na obrázku 10 vľavo je znázornený meplot pre tieto dáta. Obrázok vpravo znázorňuje ten istý meplot, ale po vypustení 4000 hodnôt. Z tohto grafu ľahko zvolíme vhodnú hranicu, od ktorej je priebeh grafu skoro lineárny. Táto hranica môže byť na úrovni 4.88. Odhadnú sa parametre GPD rozdelenia, ktoré by malo popisovať dáta nad touto hranicou. Konkrétne hodnoty parametrov sú $\hat{\beta} = 0.1769403$ a $\hat{\xi} = 0.7247035$. Počet hodnôt nad touto hranicou je 8796. Dosadením

do vzorca (33) sa dostanú príslušné kvantily. Kvantily $g - h$ rozdelenia dostaneme priamo zo vzorca:

$$F^{-1}(\alpha) = a + bk(\phi^{-1}(\alpha)),$$

$$\text{kde } k(x) = \frac{e^{gx} - 1}{g} e^{hx^2/2}.$$

Tabuľka 10 nám ukazuje jednotlivé kvantily získane pomocou metódy POT a kvantily $g - h$ rozdelenia.



Obr. 10: Graf stredných hodnôt excesov $g - h$ rozdelenia s parametrami

$$a = 5; b = 0,2; g = 1,5; h = 0,2$$

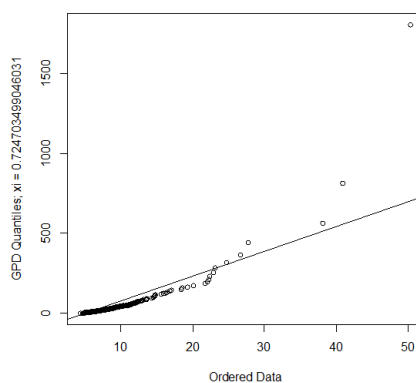
Kvantily	POT	$g - h$
90%	5,816	5,917
95%	6,586	6,886
99%	10,897	12,272
99,9%	37,855	40,344

Tabuľka 2: Kvantily $g - h$ rozdelenia a odhady kvantilov metódou POT, pri ktorej bola zvolená hranica $u = 4,88$.

Zdroj: Vlastná tvorba

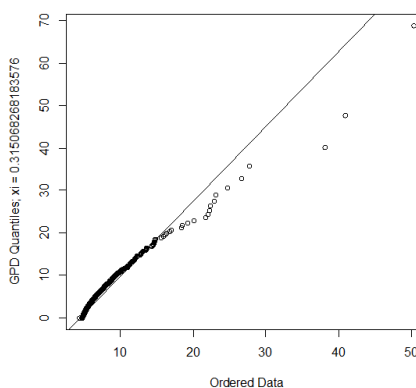
Z tabuľky 10 je vidieť, že kvantily odhadované pomocou metódy POT nie sú úplne rovnaké, ako pre $g - h$ rozdelenie. Pri 90 a 95 percentných kvantiloch výchyľky nie sú veľmi veľké, ale v prípade 99 a 99,9 percentného kvantilu je ten rozdiel väčší, čo je spôsobené hlavne voľbou vhodnej hranice.

Na obrázku 11 je zobrazený QQplot pre tieto dáta. Je vidieť isté vychýlenie.



Obrázok 11: QQplot pre dáta s hranicou 4,88

V prípade, ak zvolíme inú hranicu, napríklad 9.5, QQplot nám ukazuje lepšiu zhodu, ako je to zobrazené na obrázku 12. Avšak odhadované kvantily pomocou POT metódy pre hranicu 9.5 sa omnoho viac líšia od kvantilov $g - h$ rozdelenia, ako tomu bolo pri hranici 4,88. Tabuľka 3 porovnáva tieto kvantily.



Obrázok 12: QQplot pre dáta s hranicou 9,5

Kvantily	POT	$g - h$
90%	4,206	5,917
95%	5,426	6,886
99%	9,531	12,272
99,9%	20,534	40,344

Tabuľka 3: Kvantily $g - h$ rozdelenia a odhady kvantilov metódou POT, pri ktorej bola zvolená hranica $u = 9,5$.

Zdroj: Vlastná tvorba

Je vidieť, že voľba hranice je veľmi dôležitá a veľmi podstatná vec pri POT metóde. Pri zlej voľbe hranice môžu byť odhady kvantilov nespoľahlivé a môžu sa podstatne líšiť od kvantilov $g - h$ rozdelenia.

Záver

Cieľom bakalárskej práce bolo poukázať a priblížiť jednotlivé parametrické rozdelenia. Práca je zameraná hlavne na problematiku $g - h$ rozdelenia, jeho zavedenie, definíciu a hlavne jeho vlastnosti. Práca poukazuje na chovanie sa chvostov a ich vlastnosti pre toto rozdelenie. A keďže toto rozdelenie je rozdelenie s ťažkým chvostom, poukazuje tým na jeho využitie v poisťovníctve, hlavne pri teórii extrémnych hodnôt.

V práci bola rozobratá POT metóda a postup hľadania vhodnej hranice. Ukázalo sa, že túto metódu je vhodné použiť k odhadovaniu pravdepodobností prekročenia veľmi vysokých hraníc, alebo k odhadovaniu vysokých kvantilov

V závere bol uvedený simulačný príklad pre odhad kvantilov pomocou metódy POT a hodnoty boli porovnané s kvantilmi pre $g - h$ rozdelenie. Voľba hranice pre POT metódu je náročná, a preto odhadované kvantily môžu byť nespoľahlivé. Je to vidieť aj v príklade, kde odhadnutý 99,9%-ný kvantil metódou POT sa líši od 99,9%-ného kvantilu $g - h$ rozdelenia.

Literatúra

- 1) BIS, Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version, 2006. Dostupné z: <http://www.bis.org/publ/bcbs128.pdf>
- 2) DUTTA, K. K. a D. F. BABBEL. On Measuring Skewness and Kurtosis in Short Rate Distributions: The Case of the US Dollar London Inter Bank Offer Rates, 2002, Wharton Financial Institutions Center Working Paper.
- 3) DUTTA, K. a J. PERRY. A tale of tails: An empirical analysis of loss distribution models for estimating operational risk capital, 2007. FRB of Boston Working Paper No. 06-13. Dostupné z: <http://www.bos.frb.org/economic/wp/wp2006/wp0613.pdf>
- 4) EMBRECHTS, P., M. DEGEN a D. LAMBRIGGER. The quantitative modeling of operational risk: Between g – and $-h$ and EVT. In: DOI: 10.2143/AST.37.2.2024067, 2007. Dostupné z: ftp://ftp.sam.math.ethz.ch/hg/users/embrecht/g-and-h_May07.pdf
- 5) KLEIN, T., V. BAHÝL a V. VACEK. Základy pravdepodobnosti a matematickej štatistiky. 2. vyd. Zvolen, 2002. ISBN 80-228-1126-2.
- 6) MARKECHOVÁ, D., B. STEHLÍKOVÁ a A. TIRPÁKOVÁ. Štatistické metódy a ich aplikácie. Nitra, 2011. ISBN 978-80-8094-807-8. Dostupné z: http://www.km.fpv.ukf.sk/upload_publikacie/20120125_143707__1.pdf
- 7) MANDL, P., L. MAZUROVÁ a I. JUSTOVÁ. Matematika a řízení rizik 2009/10. 1. vyd. Praha: MATFYZPRESS, 2010. ISBN 978-80-7378-132-3.
- 8) MCNEIL, A. J. a T. SALADIN. The Peaks over Thresholds Method for Estimating High Quantiles of Loss Distributions, 1997. Proceedings of 28th International ASTIN Colloquium. Dostupné z: <http://www.math.ethz.ch/~mcneil/ftp/cairns.pdf>
- 9) SANDERS, D. The management of losses arising from extreme events, GIRO 2002. Dostupné z:

<http://www.actuaries.org.uk/sites/all/files/documents/pdf/management-losses-arising-extreme-events2002.pdf>

- 10) SIVÁK, R., Ľ. GERTLER a U. KOVÁČ. Riziká a modely vo financiách a v bankovníctve. 2.vydanie. Bratislava, 2010. Edícia Economics. ISBN 978-80-89393-44-2.
- 11) ZÁKON č. 664 zo 6. decembra 2006, ktorým sa mení a dopĺňa zákon č. 483/2001 Z. z. o bankách a o zmene a doplnení niektorých zákonov v znení neskorších predpisov a o zmene a doplnení niektorých zákonov. In: Zbierka zákonov č. 644/2006. 2006, 249. Dostupné z:
<http://www.zbierka.sk/sk/predpisy/644-2006-z-z.p-9744.pdf>